A METHOD OF CREATING A SEMANTIC VIDEO SUMMARY
USING INFORMATION FROM SECONDARY SOURCES

5   BACKGROUND OF THE INVENTION

The present invention relates to a method of summarizing or abstracting video and, more particularly, a method for using information related to video obtained from a source other than the video to create an audio-video semantic summary of video.

10   The dramatic increase in the quantity of available video, a trend which is expected to continue or accelerate, has increased the need for an automated means of summarizing video.   A summary of a video could be viewed as a preview to, or in lieu of, viewing the complete, unabridged video. Summarization could also be used as a basis for filtering large quantities of available video to

15   create a video abstraction related to a specific subject of interest. However, to be most beneficial the summary or abstraction should be semantically significant, capturing major events and meaning from the video.

There are three broad classes of techniques for creating video summaries. A first class of techniques produces a linear summary of a video sequence.   A

20   linear summary comprises a collection of key frames extracted from the video. Groups of similar frames or shots are located in the video sequence, and one or more key frames are selected from each shot to represent the content of the shot. Shot boundary detection and selection of key frames within a shot are based on lower level video analysis techniques, such as frame to frame variation in color

25   distribution or temporal positioning of a frame in a shot. While the creation of linear summaries can be automated, the extraction of a linear summary is not event driven and may only capture a rough abstraction of the video. Linear summaries are useful for video sequences where events are not well defined, such as home video, but are not well suited to producing meaningful summaries of

30   videos containing well defined events, such as videos of sporting events.

A second summary extraction technique produces a video story board. The story board is a graphic presentation of the video comprising a number of nodes and edges. Nodes are created by grouping shots usually on the basis of some low level visual characteristic, such as a color histogram. The edges

5   describe relationships between the nodes and are created by human interaction with the summarizing system. While story boarding can produce meaningful summaries, it relies on human intervention to do so.

A third summary extraction technique involves the creation of semantic video summaries which requires an understanding of the events in the video and,

10   in many cases, some expertise of the domain or subject area portrayed by the video. Obtaining this understanding and expertise through automated means has, heretofore, been problematic. Smith et al., VIDEO SKIMMING FOR QUICK BROWSING BASED ON AUDIO AND IMAGE CHARACTERIZATION, Carnegie-Mellon University Tech Report, CMU-CS-95-186, 1995, utilizes detection of

15   keywords in the audio track or closed captioning accompanying a video as a basis for locating meaningful video segments. However, it is difficult to select appropriate keywords and the selected keywords may be uttered many times as part of some general commentary related to the subject of the video without necessarily signaling the presence of corresponding meaningful visual images.

20   What is desired, therefore, is a method of creating meaningful event driven video summaries that minimizes the necessity for human intervention in the summarizing process.

SUMMARY OF THE INVENTION

25   In a first aspect, the present invention overcomes the aforementioned drawbacks of the prior art by providing a method of creating a semantic summary of a video comprising the steps of identifying a domain of the video; using the domain identity to locate information related to the video at a source other than the video; extracting data from the information; and extracting content from the

30   video related to the data. The domain or sphere of activity of the video can be

identified from an electronic programming guide or other indexing means. With the domain identity known, the summarizing system can link to a worldwide web site or an electronic programming guide containing statistics or other summary information about events of the video. In this way, the system can obtain the high

5 level knowledge about the progress of events in the video to permit the system to identify content corresponding to significant events. Thus, although the summarizing information was created for other purposes it can be used to assemble a semantic summary of the video.

In a second aspect of the invention, a method of abstracting video is

10 provided comprising the steps of locating an index of the video; identifying a domain of a video abstraction; using the domain identity to identify in the index video for inclusion in the abstraction; and extracting the identified video. An index of a video collection is located in an electronic programming guide, on the worldwide web or otherwise. When the domain of the particular abstraction is

15 identified, the domain's identity can be used as a basis for searching the collection for video relevant to the desired abstraction.

The method of the present invention provides the high level domain knowledge necessary for creating semantic summaries and abstractions by locating the knowledge for event recognition and detection in existing sources

20 without the need for additional human intervention.

The foregoing and other objectives, features and advantages of the invention will be more readily understood upon consideration of the following detailed description of the invention, taken in conjunction with the accompanying drawings.

25

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of the summarization and abstraction system of the present invention.

FIG. 2 is an exemplary worldwide web page containing information related

30 to a sporting event.

FIG. 3 is an exemplary video index containing programming information related to headline news.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

5   Creation of a semantically significant summary of video footage generally requires the ability to recognize and detect events and objects in the visual images, together with some high-level knowledge of the domain or sphere of activity of the video upon which to base the inference of the occurrence of an event in the video. For example, providing a meaningful summary of a sporting

10   event requires knowledge concerning the progress of events in the game to permit the summarizing system to detect significant plays, such as scoring attempts, and the capability to locate and extract a visual representation of the detected event. While computer vision techniques are increasingly capable of object recognition and detection, event recognition requires a higher level of understanding of the

15   progress of events than computers have been capable of achieving by processing video.

           The inability of automated techniques to develop a sufficient understanding of events in the domain of a video, in such a manner as to enable detection of events, is a significant problem in automating the creation of semantically

20   significant video summaries. While this remains a problem, the present inventors realized that there are automatically accessible sources of information separate from the video itself that index or summarize events in the domain of many videos. Further, even though this information was created for other purposes, it may be used for creating semantic summaries or abstractions of video. The method of

25   creating a semantic video summary of the present invention uses available information from sources other than the video, such as world-wide-web sites, an electronic programming guide, or other sources, to aid in event inference and detection to create meaningful summaries.

           Referring to FIG. 1, the summarization method of the present invention first

30   identifies a domain 1 or sphere of activity in which a semantic summary system 2

-4-

will operate. Higher level domains include sports, news, movies, etc. These domains may be further divided into more specific subject areas, such as basketball, soccer, football, etc. The identity of the domain of a broadcast video may be extracted from an electronic programming guide that lists television

5   programming and content for viewers. Other indexing systems related to video may also be used to identify the domain. Based on the identification of the domain, the summarizing system 2 links to a secondary source of information 4, such as a worldwide web site, which includes statistical, summary, or other indexing information related to the events which are the subject of the video 6.

10 Many worldwide web sites include statistics, summary information, or commentary related to sports activities, news, or movies. Much of the information at worldwide web sites is in the form of textual summaries. Using optical character recognition, parsing of hypertext markup language (HTML) code or other known techniques, the summarization system 2 extracts information that is semantically significant to

15 the video 6.

For example, FIG. 2 illustrates an exemplary textual summary related to a soccer game posted on a worldwide web site. From such a textual summary the video summarization system 2 can extract information, such as names and times of events, and identification of subjects (e.g., names and uniform numbers of

20 players). Using the data obtained from the summary, the summarization system 2 can pinpoint a number of shots from the video 6 around the times of the scoring events for inclusion in the video summary 8. On the other hand, the system might use object detection and optical character recognition to locate the specific players to further pinpoint video content corresponding to the events identified in

25 the textual summary described at the secondary source of information 4. The summarization system 2 permits a user to enter preferences 10 related to the summary 8 to be created. For example, a user might instruct the system to search out events related to a particular individual or subject.

The summarization system 2 can also be used to create an abstraction of a

30 video collection. Utilizing indexing information for the video collection extracted

from the secondary source of information 4 and a domain for the video abstraction 1 provided by a user, the summarization system 2 can search video footage 6 that may be stored in a home file system or filter broadcast video to identify portions of the video matching events or subjects in the requested domain.

5   For example, the summarization system 2 can extract information related to source, story title, names of subjects, or times of events from a site summarizing televised news programming.

FIG. 3 illustrates an exemplary textural index of broadcast news events. If the user wishes to review video footage related to "DuPont," a keyword search by

10 the summarization system of such a news index at a worldwide web site or in an electronic programming guide would locate two relevant news events. From this information the system 2 can direct the recording of the appropriate videos for an abstraction 8 of the news programming for eventual presentation to the viewer. Existing audio or visual information from a secondary source 4, such as a web

15 site, may also be down loaded from the secondary source 4 and matched to the video segments obtained from broadcast media.

All the references cited herein are incorporated by reference.

The terms and expressions that have been employed in the foregoing specification are used as terms of description and not of limitation, and there is no

20 intention, in the use of such terms and expressions, of excluding equivalents of the features shown and described or portions thereof, it being recognized that the scope of the invention is defined and limited only by the claims that follow.